

Interoperabilidade de repositórios digitais: Em busca da preservação digital

Jean Frederick Brito Xavier¹

¹ Doutorando do Programa de Pós-Graduação em Ciência da Informação da Universidade de São Paulo (USP), Mestre em Ciência da Informação Mestre em Ciência da Informação pela Universidade Federal do Rio de Janeiro / Instituto Brasileiro de Informação em Ciência e Tecnologia UFRJ-IBICT, Arquivista da Universidade Federal de São Paulo (UNIFESP). Email: jeanfrederick3@hotmail.com

Resumo

Introdução: Com o avanço das Tecnologias de Informação e Comunicação, nossa sociedade tem se transformado e esse processo impacta na forma de gerir e disponibilizar as informações. Os Repositórios Digitais têm se destacado por serem ferramentas que auxiliam o acesso e cada vez mais estão sendo relacionados a Preservação Digital. Esse trabalho busca levantar os pontos marcantes sobre esses dois grandes temas e se enquadra no eixo temático "Repositórios Institucionais e Temáticos: Gestão e curadoria, Interoperabilidade, Visibilidade, Preservação digital e Avaliação de repositórios". **Problema e Justificativa:** Repositórios institucionais e temáticos são parte da solução sistêmica que buscamos para garantir a preservação e o acesso autêntico das informações produzidas por nossa sociedade. Esse problema tem permeado as discussões dos profissionais da Informação, tendo em vista a necessidade de tratar a informação digital. **Objetivo:** Discorrer sobre os aspectos que relacionam os Repositórios Digitais com o campo da Preservação Digital e buscar evidenciar a convergência entre ambos. **Metodologia:** A metodologia da pesquisa exploratória foi empregada neste artigo com o objetivo de proporcionar um panorama amplo sobre o tema. Foi realizada revisão bibliográfica e levantamento de dados sobre Preservação Digital e uso de Repositórios Digitais. **Resultados:** Constata-se que os Repositórios Digitais tendem a incorporar conceitos da Preservação Digital, tal como é disposto no modelo OAIS. **Conclusões:** A solução para a preservação da informação digital não está na aplicação de técnicas isoladas, mas sim no planejamento de procedimentos e estratégias encontradas teoricamente no OAIS e na prática, no Archivemática.

Abstract

Introduction: With the advancement of Information and Communication Technologies, our society has been transformed and this process impacts on how to manage and make available the information. The Digital Repositories have stood out because they are tools that help the access and more and more are being related to Digital Preservation. This work seeks to raise the important points about these two major themes and falls within the thematic axis "Institutional and

Thematic Repositories: Management and Curatorship, Interoperability, Visibility, Digital Preservation and Repository Evaluation". **Problem and Justification:** Institutional and thematic repositories are part of the systemic solution that we seek to guarantee the preservation and authentic access of the information produced by our society. This problem has permeated the discussions of information professionals, in view of the need to treat digital information. **Objective:** To discuss the aspects that relate the Digital Repositories with the field of Digital Preservation and to seek to show the convergence between both. **Methodology:** The exploratory research methodology was used in this article with the aim of providing a broad panorama on the theme. A bibliographic review and data collection on Digital Preservation and use of Digital Repositories was carried out. **Results:** Digital repositories tend to incorporate Digital Preservation concepts, as set out in the OAIS model. **Conclusions:** The solution for the preservation of digital information is not in the application of isolated techniques, but in the planning of procedures and strategies found theoretically in OAIS and in practice in Archivemática.

Introdução

A humanidade tem produzido documentos em escala crescente desde a invenção da imprensa com Gutenberg até os dias atuais. Esse fenômeno ganhou impulso principalmente após o advento dos computadores e dos documentos digitais, que, após a segunda guerra mundial, tiveram seu uso expandido das organizações militares para a população em geral, alcançando as instituições públicas e privadas. A evolução das Tecnologias de Informação e Comunicação (TIC) resultou em uma mudança paradigmática na forma da sociedade gerenciar as suas informações. Gradativamente, com o desenvolver dessa nova dinâmica, os meios tradicionais de comunicação, como o texto impresso, passaram a ser substituídos pela tecnologia digital que proporciona a otimização do processo de produção e divulgação da informação.

A informática traz vantagens como, por exemplo, a facilidade de produção documental em larga escala. Por isso, tornou-se gradativamente mais complexa a tarefa de disponibilizar as informações, seja para o agente da administração, que produz e gerencia as informações, como também, para os pesquisadores, que buscam acessar as informações para fins culturais ou científicos. Sendo assim, após esse primeiro momento de euforia motivado pelos avanços tecnológicos foi constatada a necessidade de se buscar formas para garantir o acesso à massa informacional produzida pelos indivíduos, pelas organizações e governos.

Em resposta às novas demandas para produzir, gerir, disponibilizar e preservar as informações digitais surgem novas soluções como os Repositórios Digitais e softwares de gestão e acesso à informação digital. É importante que as diversas soluções possam interoperar e trabalhar de forma sistêmica e macro, como uma única solução. Com um papel de destaque nesta discussão estão os Repositórios Digitais, que estão evoluindo e expandindo suas funcionalidades, estreitando consideravelmente a sua relação com a Preservação Digital.

Preservação tradicional x preservação digital

Podemos encontrar em Arquivos, Bibliotecas e Museus documentos em suporte de papel, que resistiram às ações do tempo, do manuseio e do ambiente externo por séculos (muitas vezes mesmo sem nenhuma intervenção ou conservação). No entanto, no caso dos documentos digitais há o constante risco de perda de parte da informação ou, até mesmo, de toda informação contida mesmo que pouco tempo após a sua criação.

Com os documentos produzidos em suportes tradicionais como o papel, existe uma preocupação relativa à preservação, que visa a garantir o documento materializado por séculos, com o mínimo de intervenção. Busca-se sempre evitar intervir no documento e protegê-lo dos ataques de pragas e dos efeitos negativos resultantes das reações químicas, que são desencadeadas com a variação da temperatura e umidade do ar (ambiente externo ao documento).

No meio digital, sabemos que instantaneamente pode-se perder completamente todo o material produzido por uma pessoa física ou jurídica, caso não existam sistemas de resgate, *backup*, e diversos outros requisitos relacionados à segurança da informação. Isso significa que a degradação no ambiente digital pode ser instantânea, muito mais rápida do que no caso dos documentos tradicionais.

Na figura seguinte é apontada a diferença de procedimentos referentes à preservação tradicional e à preservação digital:

Figura 1: Ciclos de vida da preservação tradicional e da preservação digital



Fonte: Adaptado de Blue Ribbon Task Force On Sustainable Digital Preservation And Access (2010, p. 25)

De acordo com a Figura 1, podemos perceber que na preservação dos documentos tradicionais, a preservação é somente pensada no momento em que os documentos chegam às instituições responsáveis pela preservação, como Arquivos, Bibliotecas e Museus. Na Preservação digital podemos perceber

que as "ações de preservação" acontecem desde o momento da criação dos documentos até o momento do seu acesso a longo prazo.

Rondinelli explica a diferença da preservação de documentos em suporte de papel para documentos digitais, principalmente o que tange ao seu processo de deterioração, quando nos afirma que:

Quando se trata de documentos de papel, apesar das diversas possibilidades de perda da informação, podemos sentir um maior conforto, pois a não ser que aconteça algum sinistro, estaremos observando o processo de deterioração do papel que dependerá da qualidade do papel, do ambiente em que ele está acondicionado e diversos outros fatores. Quando tratamos de documentos digitais, nos deparamos com o desafio de tratar de documentos que a qualquer momento podem sumir completamente sem deixar grandes rastros, caso não seja implementada uma política de preservação digital dos acervos (RONDINELLI, 2011, p.55)

De acordo com Ferreira (2006), preservação digital é o conjunto de atividades ou processos responsáveis por garantir o acesso continuado a longo prazo à informação. Consiste assim, na capacidade de garantir que a informação digital permaneça acessível e com qualidades de autenticidade suficientes para que possa ser recuperada integralmente no futuro, recorrendo a uma plataforma tecnológica diferente da utilizada no momento da sua criação

Em 2010, Sayão afirmava que as principais estratégias de preservação digital, que estavam sendo praticadas e pesquisadas, tendo como foco garantir a preservação da informação digital, eram a preservação da tecnologia, a emulação, a migração e o encapsulamento. Segundo o autor esses procedimentos são explicados da seguinte forma:

a) Preservação da tecnologia: estratégia baseada na criação de museus tecnológicos que mantêm equipamentos e *software* obsoletos, de forma que os documentos digitais possam ser processados no seu ambiente original. É uma solução de curto prazo;

b) Emulação: estratégia fundamentada na premissa de que o melhor meio de preservar as funcionalidades e a aparência de um objeto informacional digital é preservá-lo junto ao seu *software* original; dessa forma, o objeto pode ser rodado em plataformas atuais por meio de emuladores, que são programas que criam mímicas do comportamento de *hardware* e sistemas operacionais obsoletos em computadores novos. Essa estratégia tem sido foco de muitas pesquisas e controvérsias;

c) Migração – tem como fundamento a migração periódica de um patamar tecnológico em vias de se tornar obsoleto e/ou de se degradar fisicamente para outro mais atualizado e íntegro, incluindo mídias, ambientes de *software*, formatos e computadores; é a estratégia correntemente mais utilizada pelas organizações (SAYÃO, 2007);

d) Encapsulamento – baseia-se na ideia de que os objetos preservados devem ser auto descritos e encapsulados em estruturas físicas ou lógicas com todas as informações necessárias para que seja decifrado e compreendido no futuro.

Além das soluções citadas, ainda são conhecidas outras soluções pontuais como a criação de museus tecnológicos, preservação da cadeia de bits e o refrescamento. Porém, na atualidade é consenso entre os teóricos da Preservação Digital, a necessidade do uso de uma solução sistêmica e que esteja atrelada a uma Política de Preservação Digital, que contemple uma resposta ampla através de uma política de preservação a ser adotada pelas Instituições

Neste contexto, os Repositórios Digitais adquirem destaque pois não só são responsáveis pela guarda da informação digital, mas pela efetiva preservação das informações digitais. Existem então dois grandes pilares que norteiam as atividades dos Repositórios para serem eficientes e confiáveis: a interoperabilidade e a Preservação Digital conforme modelo OAIS.

Um repositório digital deve estar em conformidade com as normas e padrões estabelecidos, de forma a possibilitar níveis de interoperabilidade com outros repositórios digitais e sistemas informatizados. Podem ser citados como exemplos dessas normas e padrões: o “Open Archives Initiative Protocol for Metadata Harvesting – OAI-PMH”, para coleta de registros de metadados em repositórios digitais; o “Metadata Encoding and Transmission Standard – METS”, para a codificação de metadados descritivos, administrativos e estruturais; o “Encoded Archival Description – EAD”, para a codificação de metadados descritivos de documentos arquivísticos; e os “Padrões de Interoperabilidade de Governo Eletrônico – e- PING”, no caso dos órgãos e entidades do governo federal. (BRASIL, 2015, p. 17).

A interoperabilidade como foi dito, se torna importante tendo em vista soluções que devem se integrar em um sistema único como é cada vez mais nítido na área de Arquivos, Bibliotecas, Patrimonial e inclusive de Gestão Administrativa, em que são contidas coleções heterogêneas, com padrões variáveis de metadados, mas que devem ser administrados de forma padronizada.

Sobre a Preservação Digital, que é o centro deste trabalho, destacam-se os requisitos que buscam garantir a preservação dos formatos como é estabelecido na ISO 14721:2012 - Open Archival Information System (OAIS), que é conhecido em português como SAAI (Sistema Aberto de Arquivamento de Informação). É importante destacar que apesar de ter o termo “arquivamento” ou “open archival” este é um modelo direcionado a qualquer Órgão que produza informação digital e não necessariamente a Arquivos.

Repositórios digitais X OAIS

Podemos perceber claramente em (SALES; SAYÃO, 2015) quando nos dizem que há muitos desafios para as áreas da Ciência da Informação, Biblioteconomia e Ciência da Computação em torno dos repositórios digitais de dados de pesquisa, principalmente em termos de prática biblioteconômica e de aplicação de tecnologias. As bibliotecas especializadas – cumprindo sua missão - começam a expandir e renovar suas habilidades e conhecimentos, centradas em documentos, para estabelecer as bases de uma biblioteconomia de dados

capaz de lidar com os estoques crescentes desses ativos informacionais. A seguir destacamos alguns repositórios conhecidos:

- **Fedora:** *software open source* que fornece um repositório de livre acesso por meio de ampla infraestrutura para o armazenamento, gestão e disseminação de objetos digitais, incluindo o relacionamento entre eles (Sayão; Marcondes, 2009);

- **EPrints:** *software open source*, compreendido como uma das mais usadas plataformas para repositórios institucionais. É classificado como o modo mais fácil e rápido de criar repositórios de acesso livre para base de dados científica (Sayão; Marcondes, 2009).

- **DSpace:** *software open source* que fornece um repositório com funções de captura, distribuição e preservação da produção intelectual e científica, dando visibilidade e garantindo acessibilidade no decorrer do tempo (Sayão; Marcondes, 2009).

Porém, esse trabalho busca destacar a *International Organization for Standardization ISO 14721 – Reference model for an open archival information system* (OAIS) que é o resultado do esforço do *Management Council of the Consultative Committee for Space Data Systems* (CCSDS, 2003) na produção de normas que pudessem auxiliar na preservação digital das informações decorrentes dos trabalhos relacionados às missões espaciais.

Apesar de ter sido desenvolvido para missões espaciais, o OAIS tem relevância ampla, servindo como base para discussões relacionadas ao acesso de documentos digitais e à preservação digital. O OAIS descreve os componentes funcionais que devem compor um sistema de informação focado na preservação digital, prevendo o funcionamento das interfaces internas e externas do sistema, possibilitando uma expansão das possibilidades para o uso de metadados de forma que se garanta a preservação digital.

Com o objetivo de ancorar o debate em aspectos práticos, podemos perceber a teoria do OAIS integralmente em um Repositório Digital, o Archivematica, que é um sistema de preservação digital de software livre, gratuito e de código aberto projetado para manter os dados baseados em padrões de preservação digital e o acesso a longo prazo para coleções de objetos digitais.

O aspecto que diferencia o Archivematica de outras soluções conhecidas no mercado e citados anteriormente, diz respeito principalmente ao fato de atender os requisitos relacionados a Preservação Digital constantes no OAIS.

Esta é uma tendência de todos os repositórios, se tornarem o que conhecemos como repositório digital confiável, que são repositórios capazes de manter autênticos os materiais digitais, de preservá-los e prover acesso a eles pelo tempo necessário. (BRASIL, 2015, p. 15).

Para cumprir essa missão, segundo o relatório “Trusted Digital Repositories: attributes and responsibilities” (RLG/OCLC, 2002), os repositórios digitais confiáveis devem:

- Aceitar, em nome de seus depositantes, a responsabilidade pela manutenção dos materiais digitais;

- Dispor de uma estrutura organizacional que apoie não somente a viabilidade de longo prazo dos próprios repositórios, mas também dos materiais digitais sob sua responsabilidade;
- Demonstrar sustentabilidade econômica e transparência administrativa;
- Projetar seus sistemas de acordo com convenções e padrões comumente aceitos, no sentido de assegurar, de forma contínua, a gestão, o acesso e a segurança dos materiais depositados;
- Estabelecer metodologias para avaliação dos sistemas que considerem as expectativas de confiabilidade esperadas pela comunidade;
- Considerar, para desempenhar suas responsabilidades de longo prazo, os depositários e os usuários de forma aberta e explícita;
- Dispor de políticas, práticas e desempenho que possam ser auditáveis e mensuráveis; e
- Observar os seguintes fatores relativos às responsabilidades organizacionais e de curadoria dos repositórios: escopo dos materiais depositados, gerenciamento do ciclo de vida e preservação, atuação junto a uma ampla gama de parceiros, questões legais relacionadas com a propriedade dos materiais armazenados e implicações financeiras.

Podemos concluir que um Repositório Digital não se limita a apenas uma operação mas sim um complexo sistêmico composto de normas, infraestrutura organizacional, *hardware*, *software*, metadados e procedimentos Técnicos, que visam inclusive a preservação digital para que a informação esteja acessível de forma autêntica ao usuário final.

Archivematica

O Archivematica é um software livre desenvolvido sob perspectiva da preservação digital, ao implementar procedimentos que em outra etapa eram isolados, mas agora são automatizados dentro deste software de forma sistêmica, como por exemplo a emulação e a migração dos dados, evitando a obsolescência tecnológica, a incompatibilidade de formato, versões e suportes.

Principalmente em razão da constante mudança de suportes, torna-se inviável que o profissional da Informação esteja sempre monitorando os formatos e atualizando especificamente grupos diferentes. Com o uso do Archivematica, os arquivos são automaticamente convertidos e são tratados em pacotes, tais como estabelecidos no OAIS (Pacote de Informação de Submissão – *Submission Information Package*, Pacote de Informação de Arquivamento – *Archive Information Package* e Pacote de Informação de Disseminação – *Dissemination Information Package*), facilitando assim, a gestão de grandes acervos digitais.

Os microsserviços que são realizados dentro do software, que são responsáveis pelo diferencial do mesmo diante dos outros repositórios. Tendo em vista, que o Archivematica foi baseado no *Metadata Encoding Transmission*

Protocol (METS), o mesmo gera arquivos METS para registro e consultas nos seus diretórios, assim como também por ser baseado no *Preservation Metadata: Implementation Strategies* (PREMIS) preserva os nomes dos objetos digitais nos metadados do PREMIS. O Archivematica também se baseia no padrão Dublin Core e essa flexibilidade e variedade de padrões de metadados, que auxiliaram na arquitetura do software, faz com que seu potencial de interoperabilidade seja elevado.

É importante que se destaque o microserviço de busca por vírus, que o sistema realiza, para garantir a segurança da informação, sendo assim a cada objeto digital que é inserido, o sistema realiza uma varredura para encontrar possíveis arquivos nocivos, não permitindo a contaminação do acervo.

O Archivematica mantém dois formatos de arquivos digitais: uma matriz para preservação e uma derivada para consulta. O repositório usa um sistema de segurança que a muito tempo já era utilizado em órgãos militares e de inteligência, pois possui uma “zona de segurança”, que é inacessível pelo pesquisador que esteja em uma rede. Desta forma, o usuário nunca poderia ter acesso a imagem matriz, apenas tendo acesso a imagem cópia, então mesmo que algum indivíduo tentasse invadir o sistema, no máximo teria acesso a imagem de pesquisa, sem ter qualquer entrada para a matriz, sob a segurança do repositório.

A consulta no Archivematica pode não ser satisfatória, como nos demais repositórios, porém como este repositório é um software livre, pode ser otimizado e por ser baseado nas normas já descritas, pode facilmente interoperar com outras plataformas de acesso como especialmente acontece com o ICA-ATOM ou o ATOM que são soluções criadas pela mesma Organização que criou o Archivematica, a Artefactual System. É importante destacar que este software foi resultado do esforço de diversos colaboradores como a UNESCO, o arquivo da cidade de Vancouver no Canadá, a Biblioteca da Universidade de British Columbia, o Arquivo do Rockefeller Center, dentre outros.

Conclusão

Dispomos diversas opções de Repositórios Digitais para atender nossas Instituições, cada opção com sua especificidade, porém, tendo em vista as questões levantadas inicialmente neste artigo, percebemos que é inevitável que os repositórios se atualizem para tratar com o máximo de segurança as informações sob sua custódia.

O OAIS é apontado como um modelo de referência para que seja feito realmente um arquivamento de informações digitais e não apenas uma *tag* no arquivo digital sinalizando que foi “arquivado”, sendo assim, o modelo de criação de pacotes informacionais, com metadados e todos os procedimentos citados anteriormente é apontado como o ideal na atualidade para conseguir garantir a acessibilidade e autenticidade da informação digital.

Dado a complexidade das questões atuais, a solução para a preservação da informação digital não pode ser limitada a aplicação de técnicas isoladas, mas sim na implantação de procedimentos e estratégias que encontramos teoricamente no OAIS e na prática, no Archivematica.

Referências

BLUE RIBBON TASK FORCE ON SUSTAINABLE DIGITAL PRESERVATION AND ACCESS *Sustainable economics for a digital planet: ensuring long term access to digital information*. 2010 Disponível em: <http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf> Acesso em 20 mai. 2019.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de documentos eletrônicos. Diretrizes para a presunção de autenticidade de documentos arquivísticos digitais. Rio de Janeiro: Arquivo Nacional, 2012. Disponível em: <http://www.conarq.arquivonacional.gov.br/media/diretrizes_presuncao_autenticidad_e_publicada.pdf>. Acesso em: 28 mai. 2019.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de documentos eletrônicos. Diretrizes para a implementação de repositórios digitais confiáveis de documentos arquivísticos. Rio de Janeiro: Arquivo Nacional, 2014. Disponível em: <http://www.conarq.arquivonacional.gov.br/media/publicacoes/resol_conarq_39_repositorios.pdf>. Acesso em: 13 mai. 2019.

CÉ, G; FLORES, D. Archivematica e ICA-AtoM (AtoM) como Plataformas do Ambiente de Preservação e Acesso de Documentos Arquivísticos. CENDOC, Revista: Revista Comemorativa aos 48 anos do CENDOC—Centro de Documentação da Aeronáutica. Disponível em: <https://www.academia.edu/33214169/Archivematica_e_ICA-AtoM_AtoM_como_Plataformas_do_Ambiente_de_Preserva%C3%A7%C3%A3o_e_Acesso_de_Documentos_Arquiv%C3%ADstico>. Acesso em: 09 abr. 2019.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM (CCSDS). Reference model for an open archival information system (OAIS). Blue Book. Washington, Jan. 2002. Disponível em: <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>. Acesso em: 28 mai. 2019.

FERREIRA, Miguel. **Introdução à preservação digital**: conceitos, estratégias e actuais consensos. Portugal. Universidade do Minho. Escola de Engenharia, 2006

RONDINELLI, Rosely Curi. O conceito de documento arquivístico frente à realidade digital: uma revisitação necessária – 2011. Tese (Doutorado em Ciência da Informação) – Universidade Federal Fluminense, Programa de Pós-Graduação em Ciência da Informação, Instituto de Arte e Comunicação Social, Instituto Brasileiro em Ciência e Tecnologia, Niterói, 2011.

SAYÃO, Luis Fernando. Uma outra face dos metadados: informações para a gestão da preservação digital **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, v. 15, n. 30, p.1-31, 2010.

SAYÃO Luís Fernando.; SALES L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. *Informação & Informação*, Londrina, v. 21, n. 2, p. 90-115. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939> Acesso em: 19 abr. 2019.

OCLC/RLG - Online Computer Library Center/ Research Library Group. **Preservation metadata for digital objects**: a review of the state of the art. 2001. Disponível em: <<http://www.oclc.org>>

SAYÃO Luís Fernando.; MARCONDES, Carlos Henrique. Software livres para repositórios institucionais: alguns subsídios para a seleção. In: Sayão *et al* (Org.). *Implantação e gestão de repositórios institucionais*: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009.